

OR Spectrum (2014) 36:1021–1053
 DOI 10.1007/s00291-014-0361-z

REGULAR ARTICLE

Preferential reducts and constructs in robust multiple criteria ranking and sorting

Miłosz Kadziński · Salvatore Corrente ·
 Salvatore Greco · Roman Słowiński

Published online: 28 March 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract We revisit the multiple criteria ranking and sorting methods based on ordinal regression, which accept preference information in the form of, respectively, pairwise comparisons or assignment examples for some reference alternatives. Robust ordinal regression methods consider the whole set of value functions reproducing these holistic statements provided at the input. Its impact on the recommendation is expressed in terms of the necessary and possible preference relations or assignments. We propose methods for generating explanations of this impact, showing pieces of preference information provided by the decision maker (DM), which led to the observed outcomes. In particular, the minimal set of preference information pieces, called preferential reduct, is identified to justify some result observable for the whole set of compatible value functions (e.g., the truth of the necessary relation for some pair of alternatives). Further, the maximal set of preference information pieces, called preferential construct, is discovered to reveal the conditions under which some result non-observable for the whole set of compatible value functions (e.g., the falsity of the possible relation for some pair of alternatives) is possible. Knowing such explanations, the DM can better understand the impact of each piece of preference information on the result and, in consequence, get conviction about the obtained recommendation.

M. Kadziński (✉) · R. Słowiński
 Institute of Computing Science, Poznań University of Technology, 60-965 Poznań, Poland
 e-mail: milosz.kadzinski@cs.put.poznan.pl

S. Corrente · S. Greco
 Department of Economics and Business, University of Catania, Corso Italia, 55, 95129 Catania, Italy

S. Greco
 Operations and Systems Management, University of Portsmouth, Portsmouth PO1 3DE, UK

R. Słowiński
 Systems Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland

Keywords Multiple criteria decision aiding · Additive value function · Ranking problem · Sorting problem · Explanation · Incomplete/partial information · Robust ordinal regression

1 Introduction

A recommendation that is produced in the course of a decision aiding process results from some cognitive activities aiming at exploring, interpreting, and debating about a particular decision problem. These activities are often supported by a decision support system (DSS), which is a computer-based information system intended to assist the decision maker (DM) in constructive exchange of information between the DM and the model of the decision problem, so as to get a solution being feasible and consistent with the value system of the DM. As noted in [Greer et al. \(1994\)](#), there are three main components of such a system, each playing a unique role in the decision aiding process. These are user interface, simulation model, and explanation module. The role of the user interface consists in querying the user for suitable inputs, acquiring knowledge about the problem, and presenting the output. The simulation model performs the analysis of the user's queries and generates some numerical results. Finally, the explanation module interprets these outputs and generates explanations.

These three components decide upon user acceptance of both a DSS and the recommendation arrived with its use. In particular, the required input information should be as simple as possible, and collected in a way that makes the DM convinced that her/his needs, preferences and value (s)he seeks are taken into account when working out a recommendation. Furthermore, the method itself should guarantee that the DM not only learns about the problem, but also that (s)he is convinced about the psychological convergence of the process and relative advantage of the indicated solution ([Belton et al. 2008](#)). Finally, the explanations provided by the system should justify that the recommendation is logical, valid, and correct. As noted by [Kass and Finin \(1988\)](#), such explanations prove to be useful in the development and maintenance of DSSs by making explicit the experts logic and assumptions.

Collecting the input information When it comes to collecting input information for the decision problem (concerning criteria, alternatives, evaluations, and, in particular, the DM's preferences), the effort of making DSSs user friendly has been implemented in different ways. Some of them support the DMs in structuring, decomposition and recomposition of preferences between alternatives by eliciting all required inputs by means of a dialogue conducted in the natural language [see, e.g., multi-attribute utility decomposition ([Humphreys and Wisudha 1981](#))]. Some other systems are based on a verbal description of decision-making problems [see, e.g., verbal decision analysis ([Larichev 2001](#))]; the need for this stems from the experimental results revealing that 2/3 of DMs prefer to give the evaluations and preferences in a verbal form. By verbal, we mean qualitative rather than quantitative, and ordinal rather than cardinal. In the recent years, an increasing interest has been also assigned to ordinal regression methods [see, e.g., UTA ([Jacquet-Lagrez and Siskos 2001](#))]. They expect the DMs to specify some examples of holistic judgments (e.g., pairwise comparisons or assignments to classes) on a subset of reference alternatives, rather than direct parameter values used

in the preference model. Preference information given in terms of decision examples is called indirect. Then, the DM's preference model leading to the same decision as the provided one is inferred. Such a compatible model is applied on the whole set of alternatives to arrive at a recommendation. Decision aiding methods based on the indirect preference information require relatively less cognitive effort from the DMs, being consistent with the "posterior rationality" (March 1978) and with the learning from examples used in artificial intelligence approaches (Furnkranz and Hullermeier 2010).

Arriving at a recommendation As far as the way of translating the DM's preferences to a recommendation is concerned, the decision aiding methods need to be transparent, i.e., they should clearly reveal the correspondence between the elicited information and the final evaluations of alternatives. Let us mention the Even Swaps approach (Hammond et al. 1998) as an exemplary implementation of such a transparency. An even swap is interpreted as a value trade-off, where a change of performance on one criterion is compensated by a comparable change on some other criterion, and thus a new alternative with these revised performances is equally preferred to the initial one. The DM is required to list her/his objectives explicitly by providing a sequence of swaps which are acceptable for her/him. As a result, the method provides a practical means for making trade-offs among any set of objectives across a range of alternatives. By simplifying the mechanical elements of trade-offs, it leads to the final decision through a sequence of choices supposed to be intuitive for the DM.

From another perspective, the decision rule approach (Greco et al. 2004; Slowinski et al. 2009), represents the preferences in terms of "if . . . then" decision rules such as, e.g., "if the maximum speed of car x is at least 175 km/h and its price is at most \$12,000, then car x is comprehensively at least medium". Thus, apart from requiring the DM to provide information in the form of examples of decisions, this approach employs a very natural and easily interpretable way of formulating conclusions. Each decision rule is a straightforward scenario relating evaluation of an alternative on selected criteria and its class assignment.

When it comes to ordinal regression methods, there is typically more than a single preference model instance compatible with the holistic statements. To avoid arbitrary selection of a single compatible instance, robust ordinal regression (ROR) has been proposed (Corrente et al. 2013; Greco et al. 2008). It allows taking into account all compatible preference model instances, and provides the DM with two results, the necessary and possible recommendation for the set of considered alternatives. The necessary recommendation is supported by all compatible instances of the preference model, while the possible recommendation is supported by at least one of them. In this way, the DM is forced to confront her/his value system with the results of applying the inferred model on the set of alternatives.

Explaining the recommendation Explanation of the recommended decision requires construction of arguments through which the DM will convince herself/himself and other stakeholders about its logic and validity. The issue of generating explanation of the outcome of a decision model has motivated a great variety of studies. In particular, Papamichail and French (2003) present a general framework for developing explanation facilities in evaluation systems with the use of natural language generation techniques. Ouerdane et al. (2010) aim at adapting argumentation-based approach for

providing the reasons underlying the chosen course of action. Apart from explaining the motivations, goals or results, such a mix of argumentation theory and decision aiding allows taking advantage of the expressive nature of the language used as well as handling incomplete or even contradictory pieces of information. Labreuche (2011) introduces a general framework for explaining the results of multi-attribute value theory. It is focused on the selection of arguments to be presented in the case where the decision model is based on the weights assigned to criteria. Some other representative works on providing convincing explanations to accompany recommendation can be found in Amgoud and Prade (2009), Carenini and Moore (2006), Klein (1994), and Labreuche et al. (2011).

When it comes to justifying the outcomes of ROR, Labreuche et al. (2012) investigate whether Even Swaps can be generalized to construct convincing explanations of the necessary preference relation (Hammond et al. 1998). A sequence of such swaps can be seen as the reasoning steps allowing to highlight why alternative a is necessarily preferred to alternative b , or, in general, why alternative a is the best choice. The authors set up a framework for providing such minimal explanations, proving both its practical usefulness in terms of binary domains considered for the value of criteria as well as questionability of the technique in the absence of any restriction on the size of the criteria domain. In the latter case, the sequence of preference-swaps may not be bounded. On the contrary, Greco et al. (2013) applies decision rules to explain the necessary and possible preference relations. These rules aim at explaining preference relations in terms of a minimal conjunction of elementary conditions concerning evaluations on particular criteria that imply their truth. For example, the preference, either necessary or possible, of alternative a over alternative b can be supported by a decision rule indicating the role of a strong preference on criterion g_1 and at least mild preference on criterion g_2 . In this way, the DM is provided with arguments on the role of particular criteria in the decision model, which enables her/him to better understand the conditions for the necessary and possible relations. It is worth noting that due to iterative way of working with ROR methods, the DM can easily see the consequence of an increment of preference information on the necessary and possible recommendations. This observation stimulates elicitation of new preference information or increases the DM's conviction that the current recommendation is the best.

Aim of the paper As discussed above, the use of ROR methods is intuitive for both the required indirect preference information and exhibition of a set of results derived from applying the whole set of compatible preference model instances. In this paper, we reconsider the task of generating explanations of the outcomes of ROR methods designed for supporting multiple criteria ranking and sorting problems. Without loss of generality, we focus on an additive value function preference model which is of particular interest in multiple criteria decision aiding (MCDA) because of an intuitive interpretation of numerical scores of alternatives and straightforward translation of pieces of preference information to the final result (see, e.g., Doumpos 2012; Lahdelma and Salminen 2012). Contrary to Greco et al. (2013) and Labreuche et al. (2012), the constructed explanations do not refer to the evaluations of alternatives but rather directly to preference information provided by the DM. Even if the required preference information is easily definable, like a set of pairwise comparisons or assignment examples,

it is processed in a way preventing the DM from seeing the exact relations between the provided preference information and the obtained recommendation. This is particularly true if the set of exemplary judgments is relatively great. Thus, analyzing the ranking recommendation one could ask the following reasonable questions:

- in case alternative a is necessarily preferred to alternative b , what are the minimal sets of pairwise comparisons provided by the DM that induce a necessary preference for a pair (a, b) in consequence of using all compatible value functions;
- in case alternative a is not even possibly preferred to alternative b , what are the maximal sets of the provided pairwise comparisons that admit a possible preference relation for a pair (a, b) in consequence of using at least one compatible value function;
- in case alternative a is ranked at positions between $P^*(a)$ and $P_*(a)$ ($P^*(a) \leq P_*(a)$) by all compatible value functions, what are the minimal sets of the provided pairwise comparisons that imply such a ranking interval, and what are the maximal sets of provided pairwise comparisons that admit a being ranked higher than $P^*(a)$ or lower than $P_*(a)$.

Furthermore, in the context of sorting problem, the following analogous questions are relevant:

- in case alternative a is necessarily assigned to class C_h , what are the minimal sets of assignment examples provided by the DM that induce such a necessary assignment;
- in case alternative a is possibly assigned to an interval of classes $[C_L, C_R]$, what are the minimal sets of the provided assignment examples that imply such a possible assignment, and what are the maximal sets of assignment examples that admit a being assigned to a class worse than C_L or better than C_R .

Answering these questions facilitates understanding of the correspondence between the output of the preference model and the revealed preferences, making these relations more transparent and intelligible for the DM. The arguments are formulated with respect to the previously provided exemplary judgments, being in this way consistent with intuitive reasoning of the DMs, and not forcing them to reason in terms of some abstract parameters of the model or evaluations of the alternatives. From such explanations, the DM learns about the impact of particular pieces of original preference information or their subsets, and gains insights into her/his preferences, which enhance the trust in the obtained recommendation. This confrontation makes the DM more conscious about the conditions for the necessary and possible outputs of ROR, leading her/him also to better understand the employed model. Moreover, the provided explanations constitute a starting point for a more thorough analysis. In case the DM does not accept some implications between the provided input and the obtained recommendation, (s)he can remove or revise some pieces of preference information.

The organization of the paper is the following. In the next section, we introduce basic concepts and notation that will be used in the paper. In Sect. 3, we remind existing multiple criteria ranking and sorting methods based on ROR, and we define the models that we refer to in our approach. In Sect. 4, we discuss how to compute explanations of the obtained necessary and possible outcomes in terms of the DM's preference information. For the purpose of illustration, we will consider the problem of

evaluating democracy regimes of different countries. We will refer to it in the following sections to illustrate different aspects of the proposed framework. In Sect. 5, we focus on the computational cost of the proposed procedures. In Sect. 6, we discuss how to use the provided explanations within the constructive approach to decision aiding. The last section concludes the paper.

2 Notation

We shall use the following notation:

- $A = \{a_1, a_2, \dots, a_i, \dots, a_n\}$ —a finite set of n alternatives;
- $A^R = \{a^*, b^*, \dots\}$ —a finite set of reference alternatives, on which the DM accepts to express holistic preferences; we assume that $A^R \subseteq A$;
- $G = \{g_1, g_2, \dots, g_j, \dots, g_m\}$ —a finite set of m evaluation criteria, $g_j : A \rightarrow \mathbb{R}$ for all $j \in J = \{1, 2, \dots, m\}$; although real-coded they may have ordinal or cardinal scales;
- $C_h, h = 1, \dots, p$ —pre-defined preference ordered classes such that C_{h+1} is preferred to $C_h, h = 1, \dots, p - 1$; $H = \{1, 2, \dots, p\}$;
- $X_j = \{x_j \in \mathbb{R} : g_j(a_i) = x_j, a_i \in A\}$ —the set of all different evaluations on $g_j, j \in J$; without loss of generality, we assume that all criteria have an increasing direction of preference, i.e., the greater $g_j(a_i)$, the more preferred alternative a_i on criterion g_j , for all $j \in J$ and $a_i \in A$;
- $x_j^1, x_j^2, \dots, x_j^{n_j(A)}$ —the ordered values of $X_j, x_j^k < x_j^{k+1}, k = 1, \dots, n_j(A) - 1$, where $n_j(A) = |X_j|$ and $n_j(A) \leq n$; consequently, $X = \prod_{j=1}^m X_j$ is the evaluation space.

To represent DM's preferences, we shall use a model in the form of an additive value function:

$$U(a) = \sum_{j=1}^m u_j(g_j(a)) \quad (1)$$

where the marginal value functions $u_j, j \in J$, are monotone, non-decreasing and their sum is normalized, so that the additive value (1) is bounded within the interval $[0, 1]$. Note that for simplicity of notation, one can write $u_j(a), j \in J$, instead of $u_j(g_j(a))$. Consequently, the basic set of constraints E_{BASE}^{AR} defining conditions on the monotonicity and normalization of general additive value functions has the following form:

$$\left. \begin{aligned} [B1] \quad & u_j(x_j^k) - u_j(x_j^{(k-1)}) \geq 0, \quad k = 2, \dots, n_j(A), \quad j = 1, \dots, m, \\ [B2] \quad & u_j(x_j^1) = 0, \quad j = 1, \dots, m, \quad \sum_{j=1}^m u_j(x_j^{n_j(A)}) = 1. \end{aligned} \right\} E_{\text{BASE}}^{AR}$$

Defined in this way, the general non-linear monotonic marginal value functions, with all criteria values being their characteristic points, do not involve any arbitrary or restrictive parametrization. However, the selection of the form of marginal value functions involves a trade-off between their flexibility and simplicity. Thus, instead of

general marginal value functions, we can use piecewise linear or even linear ones. Then, for each u_j , $j = 1, \dots, m$, we need to define the number of characteristic points γ_j . The intervals $[x_j^1, x_j^{n_j(A)}]$ are divided into $\gamma_j - 1$ equal sub-intervals with the endpoints:

$$g_j^s = x_j^1 + \left(x_j^{n_j(A)} - x_j^1 \right) (s - 1) / (\gamma_j - 1), \quad s = 1, \dots, \gamma_j.$$

Then, the monotonicity constraints [B1] should be formulated in the following way:

$$\left\{ \begin{array}{l} u_j(g_j^s) - u_j(g_j^{s-1}) \geq 0, \quad j = 1, \dots, m, \quad s = 1, \dots, \gamma_j, \\ u_j(x_j^k) = u_j(g_j^{s-1}) + \left(u_j(g_j^s) - u_j(g_j^{s-1}) \right) (x_j^k - g_j^{s-1}) / (g_j^s - g_j^{s-1}) \\ \text{for all } x_j^k \in [g_j^{s-1}, g_j^s], \quad j = 1, \dots, m, \quad k = 1, \dots, n_j(A). \end{array} \right. \quad (2)$$

3 Robust ordinal regression based on multi-attribute value theory

The majority of MCDA methods require specification of precise numerical values for parameters of the preference model. It is not always realistic to assume that the DM can provide exact values for all these parameters. Thus, disaggregation procedures that enable model inference with exemplary judgments on a set of reference alternatives have been proposed. However, usually there is more than a single model compatible with the given holistic judgments. The traditional disaggregation approaches force selecting a single such model for working out a recommendation. Such a choice is rather arbitrary, requiring either the use of some pre-defined rules or participation of the DM. Instead of defining or selecting the single model, ROR enables analysis of the whole set of compatible preference model instances. Examination of its application on the set of alternatives A leads to identifying the necessary and possible consequences confirmed by all or at least one compatible preference model instance, respectively.

3.1 Robust ordinal regression for multiple criteria ranking

As far as methods designed for dealing with multiple criteria ranking problems are concerned, ROR has been implemented for the first time in UTA^{GMS} (Greco et al. 2008). It enables the DM to provide pairwise comparisons for some reference alternatives. The comparison of a pair $(a^*, b^*) \in B^R \subseteq A^R \times A^R$ may state the strict preference, weak preference, or indifference denoted by $a^* \succ_{DM} b^*$, $a^* \succeq_{DM} b^*$, and $a^* \sim_{DM} b^*$, respectively. Let each pairwise comparison from B^R be denoted by B_k^R , $k = 1, \dots, |B^R|$. The set of constraints $E_{PI-RANK}^{A^R}$ given below translates such a reference pre-order provided by the DM to a value function:

$$\left. \begin{array}{ll} U(a^*) \geq U(b^*) + \varepsilon & \text{if } a^* \succ_{DM} b^*, \\ U(a^*) = U(b^*) & \text{if } a^* \sim_{DM} b^*, \\ U(a^*) \geq U(b^*) & \text{if } a^* \succeq_{DM} b^*, \end{array} \right\} B_k^R, \quad k = 1, \dots, |B^R| \quad \left. \vphantom{\begin{array}{l} U(a^*) \geq U(b^*) + \varepsilon \\ U(a^*) = U(b^*) \\ U(a^*) \geq U(b^*) \end{array}} \right\} E_{PI-RANK}^{A^R}$$

where ε is an arbitrarily small positive value.

The pairwise comparisons provided by the DM form the input data for the ordinal regression that finds the whole set of value functions being able to reconstruct these judgments. Such value functions are compatible with the preference information. Precisely, a set of such compatible value functions $\mathcal{U}_{\text{RANK}}^{A^R}$ is defined by a set of constraints $E_{\text{RANK}}^{A^R}$ given below:

$$\left. \begin{array}{l} E_{\text{PI-RANK}}^{A^R} \\ E_{\text{BASE}}^{A^R} \end{array} \right\} E_{\text{RANK}}^{A^R} \quad (3)$$

Thus, compatible value functions need to be monotonic and normalized (see $E_{\text{BASE}}^{A^R}$) and reproduce DM's pairwise comparisons (see $E_{\text{PI-RANK}}^{A^R}$). If $\varepsilon^* = \max \varepsilon$, s.t. $E_{\text{RANK}}^{A^R}$ is >0 and $E_{\text{RANK}}^{A^R}$ is feasible, the set of compatible value functions is non-empty. Otherwise, the provided preference information is inconsistent with the assumed preference model.

Necessary and possible preference relations UTA^{GMS} applies all compatible value functions $\mathcal{U}_{\text{RANK}}^{A^R}$, and defines two binary relations in the set of alternatives A :

- Necessary weak preference relation, \succsim^N , that holds for a pair of alternatives $(a, b) \in A \times A$, in case $U(a) \geq U(b)$ for all compatible value functions;
- Possible weak preference relation, \succsim^P , that holds for a pair of alternatives $(a, b) \in A \times A$, in case $U(a) \geq U(b)$ for at least one compatible value function.

The following linear programs need to be solved to assess whether the relations hold:

$$\text{Maximize : } \varepsilon \quad (4)$$

s.t.

$$\left. \begin{array}{l} U(b) - U(a) \geq \varepsilon, \\ E_{\text{RANK}}^{A^R} \end{array} \right\} E^N(a, b)$$

and

$$\text{Maximize : } \varepsilon \quad (5)$$

s.t.

$$\left. \begin{array}{l} U(a) - U(b) \geq 0, \\ E_{\text{RANK}}^{A^R} \end{array} \right\} E^P(a, b)$$

Then, $a \succsim^N b$ if $E^N(a, b)$ is infeasible or $\varepsilon_* = \max \varepsilon$, s.t. $E^N(a, b)$, is not >0 . $a \succsim^P b$ if $E^P(a, b)$ is feasible and $\varepsilon^* = \max \varepsilon$, s.t. $E^P(a, b)$, is >0 .

Extreme ranks Kadziński et al. (2012) extend UTA^{GMS} with the framework of extreme ranking analysis (ERA) to consider all complete pre-orders compatible with the preference information and to determine the best $P^*(a)$ and the worst $P_*(a)$

ranks taken by each alternative $a \in A$. Identification of these extreme ranks requires consideration of the following mixed-integer linear programming (MILP) models:

$$\text{Minimize : } f_{\max}^{\text{rank}} = \sum_{b \in A \setminus \{a\}} v_b \quad (6)$$

s.t.

$$\left. \begin{array}{l} U(a) - U(b) + Mv_b \geq 0, \quad \text{for all } b \in A \setminus \{a\}, \\ E_{\text{RANK}}^{A^R}, \end{array} \right\} E_{\max}^{A^R}(a),$$

and

$$\text{Minimize : } f_{\min}^{\text{rank}} = \sum_{b \in A \setminus \{a\}} v_b \quad (7)$$

s.t.

$$\left. \begin{array}{l} U(b) - U(a) + Mv_b \geq \varepsilon, \quad \text{for all } b \in A \setminus \{a\}, \\ E_{\text{RANK}}^{A^R}, \end{array} \right\} E_{\min}^{A^R}(a),$$

where M and ε are auxiliary variables equal to, respectively, big and small positive values, and v_b is a binary variable ($v_b \in \{0, 1\}$) associated with comparison of a to alternative $b \in A \setminus \{a\}$. The best rank $P^*(a)$ of alternative a is $f_{\max}^{\text{rank}} + 1$ and the worst rank $P_*(a)$ is $n - f_{\min}^{\text{rank}}$.

Kadzinski and Tervonen (2013) prove that if we limit $\mathcal{U}_{\text{RANK}}^{A^R}$ to exclude value functions for which two alternatives obtain an equal rank, there are no rank jumps in the interval $[P^*(a_i), P_*(a_i)]$, i.e., if $\exists U', U'' \in \mathcal{U}_{\text{ROR}}^{A^R} : \text{rank}(U', a_i) = \rho, \text{rank}(U'', a_i) = \rho + \delta^\rho$ with $\delta^\rho \geq 2 \implies \forall \delta \in \{0, \dots, \delta^\rho\} \exists U^* \in \mathcal{U}_{\text{ROR}}^{A^R} : \text{rank}(U^*, a_i) = \rho + \delta$.

In any case, combining UTA^{GMS} and ERA, multiple criteria ranking can be approached with a robustness analysis concerning both the outcome of separate $n \cdot (n - 1)$ pairwise comparisons as well as the result of collating each of the n alternatives with all the remaining ones simultaneously.

Case study We reconsider the problem of evaluating democracy of different countries based on the following five criteria proposed by the Economist Intelligence Unit (EIU 2010): electoral process and pluralism (g_1), the functioning of government (g_2), political participation (g_3), political culture (g_4), and civil liberties (g_5). We rank 24 countries from Australasia as well as Southeast, Central and East Asia. Their evaluations are provided in Table 1.

We assume a DM to have provided preference information in the form of five pairwise comparisons of reference alternatives (see Table 2). These are consistent with respect to the considered additive model. Let us emphasize that all experimental results presented in the paper follow the use of linear marginal value functions for all criteria.

Table 1 Performance matrix for the problem of evaluating countries on the type of democracy regime

	Country	g_1	g_2	g_3	g_4	g_5
1	Australia (AU)	10.00	8.93	7.78	8.75	10.00
2	China (CN)	0.00	4.64	2.78	6.25	1.18
3	Fiji (FJ)	6.50	5.21	3.33	5.00	8.24
4	Indonesia (ID)	6.92	7.14	5.00	6.25	6.76
5	Japan (JP)	9.17	7.86	5.56	8.75	9.41
6	Kazakhstan (KZ)	2.67	2.14	3.33	4.38	5.59
7	Kyrgyzstan (KG)	5.75	1.86	2.78	5.00	5.00
8	Laos (LA)	0.00	3.21	1.11	5.00	1.18
9	Malayasia (MY)	6.08	5.71	4.44	7.50	6.18
10	Mongolia (MN)	9.17	6.08	3.89	5.63	8.24
11	Myanmar (MM)	0.00	1.79	0.56	5.63	0.88
12	New Zealand (NZ)	10.00	8.57	8.33	8.13	10.00
13	North Korea (KP)	0.83	2.50	0.56	1.25	0.00
14	Papua New G. (PG)	7.33	6.43	4.44	6.25	8.24
15	Philippines (PH)	9.17	5.36	5.00	3.75	9.12
16	Singapore (SG)	4.33	7.50	2.78	7.50	7.35
17	South Korea (KR)	9.58	7.14	7.22	7.50	7.94
18	Taiwan (TW)	9.58	7.50	6.67	5.63	9.71
19	Tajikistan (TJ)	1.83	0.79	2.22	6.28	1.18
20	Thailand (TH)	4.83	6.43	5.00	5.63	6.47
21	Timor Leste (TL)	7.00	5.57	5.00	6.25	8.24
22	Turkmenistan (TM)	0.00	0.79	2.78	5.00	0.59
23	Uzbekistan (UZ)	0.08	0.79	2.78	5.00	0.59
24	Vietnam (VN)	0.83	4.29	2.78	4.38	1.47

Table 2 Pairwise comparisons B^R provided by the DM for the problem of ranking countries based on their democracy evaluation

$B_1^R = (JP \succ KR)$	$B_2^R = (FJ \succ TH)$	$B_3^R = (PH \succ MN)$
$B_4^R = (ID \succ MY)$	$B_5^R = (KG \succ KZ)$	

The Hasse diagram of the necessary relation \succsim^N is presented in Fig. 1. This relation is robust with respect to the given preference information meaning that a pair of alternatives is compared in the same way by all compatible value functions. Obviously, the necessary relation reproduces all pairwise comparisons of reference alternatives. Let us also remind that the necessary preference relation is transitive, and thus the arrows that can be obtained by transitivity are not represented in the Hasse diagram. Moreover, if there was no arrow representing the necessary relation between two alternatives, then these alternatives are incomparable in terms of the necessary preference.

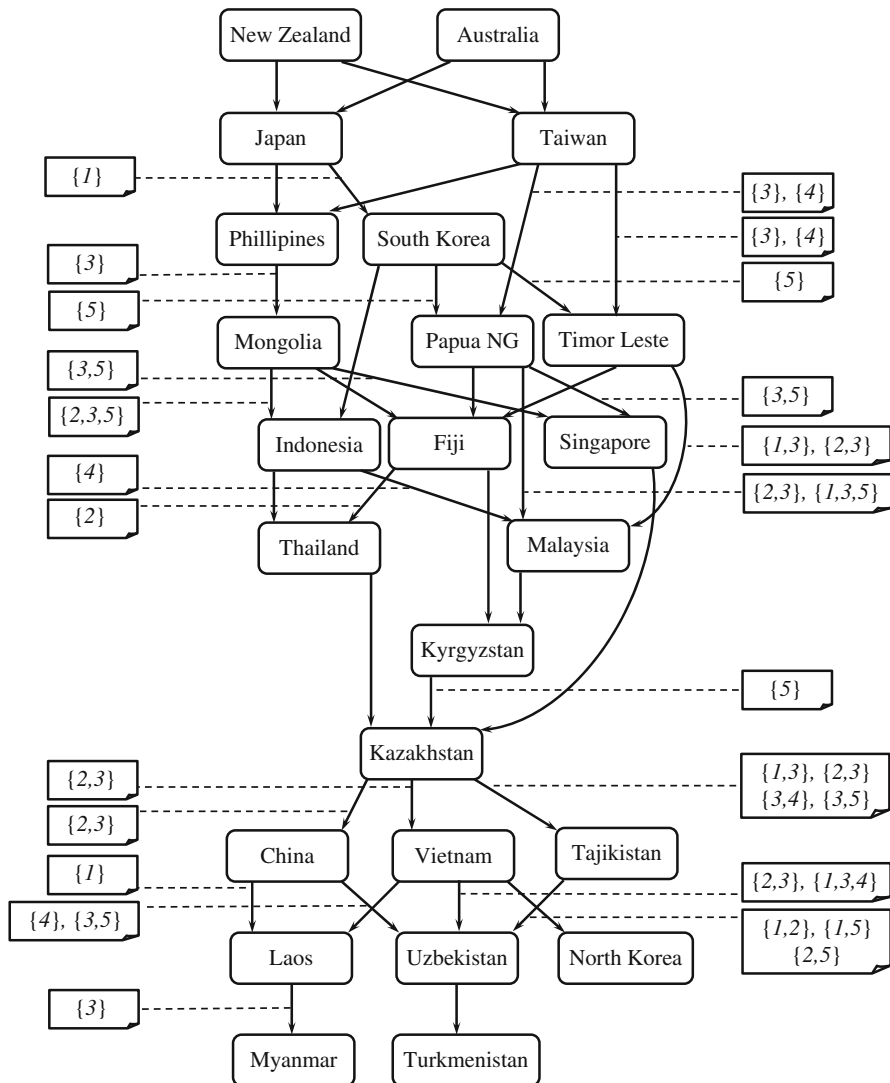


Fig. 1 Hasse diagram of the necessary relation. All preferential reducts (see Sect. 4.1.1) for the necessary relations visible in the Hasse diagram (note that we refer to indices of pairwise comparisons provided by the DM from Table 2). In case the preferential reduct is not listed, the necessary relation corresponds to the dominance relation

Table 3 shows the best $P^*(a)$ and the worst $P_*(a)$ ranks of each alternative in the context of all compatible value functions. New Zealand (NZ) and Australia (AU) should be perceived as countries with the most developed democracy, whereas Myanmar (MM), Turkmenistan (TM), and North Korea (KP) need to be considered as countries with the lowest democracy level being possibly ranked at the very bottom. One can also observe what is concordant with intuition: the alternatives which are

Table 3 Results of extreme ranking analysis for all compatible value functions ($P^*(a) - P_*(a)$) and in case no preference information is available ($P_D^*(a) - P_*^D(a)$)

Country	$P^*(a) - P_*(a)$	$P_D^*(a) - P_*^D(a)$	RPRs	RCORE
AU	1–2	1–2	–	–
CN	17–20	10–21	{1, 2, 3}	{1, 2, 3}
FJ	10–13	9–16	{2, 5}	{2, 5}
ID	8–12	5–13	{2, 3, 4}, {2, 3, 5}	{2, 3}
JP	3–4	3–6	{1}	{1}
KZ	16–16	14–23	{2, 3, 5}	{2, 3, 5}
KG	13–15	13–21	{2, 5}	{2, 5}
LA	19–23	15–24	{1, 2}, {1, 3}, {2, 3}, {3, 5}	–
MY	11–14	5–14	{2, 3, 4}, {2, 3, 5}	{2, 3}
MN	6–9	4–14	{2, 3, 5}	{2, 3, 5}
MM	20–24	15–24	{3}, {2, 4}	–
NZ	1–2	1–2	–	–
KP	19–24	18–24	{2}, {3}	–
PG	6–9	6–13	{1, 2}, {1, 3}	{1}
PH	5–8	5–23	{2, 3, 5+}	{2, 3, 5}
SG	9–15	4–16	{1, 3, 5}, {2, 3, 5}	{3, 5}
KR	4–7	3–10	{1, 5}	{1, 5}
TW	3–5	3–12	{3}	{3}
TJ	17–22	10–24	{2, 3}, {1, 3, 5}	{3}
TH	12–15	7–15	{2}	{2}
TL	6–11	4–13	{2, 3}, {3, 5}	{3}
TM	21–24	18–24	{1, 4}, {2, 3}, {2, 4}	–
UZ	20–23	18–23	{1, 3}, {1, 4}	{1}
VN	17–19	15–23	{2, 3}	{2, 3}

All rank preferential reduces (RPRs) for each alternative and the rank preferential core (RCORE) (see Sect. 4.1.2; note that we refer to indices of pairwise comparisons provided by the DM from Table 2)

good on some criteria while being relatively bad on the others can be attributed a wide range of possible ranks [see, e.g., Singapore (SG) and Timor-Leste (TL)]. On the contrary, for the alternatives whose evaluation profiles are more typical (at least within some significant subset of alternatives), the highest and the lowest possible ranks are usually close to each other [see, e.g., Australia (AU), Kazakhstan (KZ), and Taiwan (TW)].

3.2 Robust ordinal regression for multiple criteria sorting

When it comes to value-based methods designed for dealing with multiple criteria sorting problems, ROR has been implemented for the first time in UTADIS^{GMS} (Greco et al. 2010). Let us focus on a threshold-based sorting procedure, where the limits

between consecutive classes $C_h, h = 1, \dots, p$, are defined by a vector of thresholds $\mathbf{t} = \{t_1, \dots, t_{p-1}\}$ such that $0 < t_1 < \dots < t_{p-1} < 1$, and t_{h-1} and t_h are, respectively, the lower and upper threshold of class $C_h, h = 2, \dots, p-1$. Note that t_1 is an upper threshold of class C_1 while the lower threshold is 0, and t_{p-1} is a lower threshold of class C_p while there is no upper threshold t_p for C_p (or, from another perspective, it is >1 , which means that comprehensive values of all alternatives are $< t_p$).

UTADIS^{GMS} represents the DM preferences with a pair (U, \mathbf{t}) , where U is an additive value function and \mathbf{t} is a vector of thresholds delimiting the classes. It enables the DM to provide possibly imprecise assignment examples consisting of a reference alternative $a^* \in A^R$ and its desired assignment $[L_{DM}(a^*), R_{DM}(a^*)]$, with $L_{DM}(a^*) \leq R_{DM}(a^*)$. Let each assignment example be denoted by $A_k^R, k = 1, \dots, |A^R|$. These assignment examples are translated to the following constraints E_{SORT}^{AR} :

$$\left. \begin{array}{l} U(a^*) \geq t_{L_{DM}(a^*)-1}, \\ U(a^*) + \varepsilon \leq t_{R_{DM}(a^*)}, \\ t_1 \geq \varepsilon, t_{p-1} \leq 1 - \varepsilon, \\ t_h - t_{h-1} \geq \varepsilon, h = 2, \dots, p-1, \\ E_{BASE}^{AR} \end{array} \right\} \left\{ \begin{array}{l} A_k^R, k = 1, \dots, |A^R| \\ E_{PI-SORT}^{AR} \\ E_{BASE-SORT}^{AR} \end{array} \right\} E_{SORT}^{AR}$$

The set of pairs $(U, \mathbf{t})_{SORT}^{AR}$ compatible with the provided assignment examples is non-empty, if $\varepsilon^* = \max \varepsilon$, s.t. E_{SORT}^{AR} , is >0 and E_{SORT}^{AR} is feasible.

Necessary and possible assignments Given set A^R of assignment examples and a corresponding set of compatible pairs $(U, \mathbf{t})_{SORT}^{AR}$, for each alternative $a \in A$:

- The possible assignment $C_P(a)$ is defined as the set of indices of classes C_h for which there exists at least one compatible pair (U, \mathbf{t}) assigning a to C_h (denoted by $a \rightarrow^P C_h$);
- The necessary assignment $C_N(a)$ is defined as the set of indices of classes C_h for which all compatible pairs (U, \mathbf{t}) assign a to C_h (denoted by $a \rightarrow^N C_h$).

The possible assignment of $a \in A$ is composed of $h \in H$, such that $E(a \rightarrow^P C_h)$ given below is feasible and $\varepsilon^* = \max \varepsilon$, s.t. $E(a \rightarrow^P C_h)$, is >0 (Greco et al. 2010).

$$\left. \begin{array}{l} U(a) \geq t_{h-1}, \quad \text{if } h \geq 1, \\ U(a) + \varepsilon \leq t_h, \quad \text{if } h \leq p-1, \\ E_{SORT}^{AR} \end{array} \right\} E(a \rightarrow^P C_h)$$

Let us define $L_P(a)$ and $R_P(a)$ as indices of the worst and the best classes to which alternative a is possibly assigned.

The necessary assignment of alternative a is composed of $h \in H$, such that $E(a \rightarrow^N C_h)$ given below is infeasible or $\varepsilon^* = \max \varepsilon$, s.t. $E(a \rightarrow^N C_h)$, is not >0 (Kadzinski and Tervonen 2013).

Table 4 Assignment examples A^R provided by the DM for the problem of assigning countries to different types of democracy regimes

$A_1^R = (JP \rightarrow C_4)$	$A_2^R = (KR \rightarrow C_4)$	$A_3^R = (TW \rightarrow C_3)$	$A_4^R = (TH \rightarrow C_3)$
$A_5^R = (SG \rightarrow C_3)$	$A_6^R = (KZ \rightarrow C_2)$	$A_7^R = (KG \rightarrow C_2)$	$A_8^R = (CN \rightarrow C_1)$

$$\left. \begin{array}{ll}
 U(a) + \varepsilon \leq t_{h-1} + M \cdot v_1, & \text{if } h \geq 1, \\
 U(a) \geq t_h - M \cdot v_2, & \text{if } h \leq p - 1, \\
 v_1 + v_2 = 1, & \text{if } 1 \leq h \leq p - 1, \\
 v_1, v_2 \in \{0, 1\}, & \\
 E_{\text{SORT}}^{A^R}, &
 \end{array} \right\} E(a \rightarrow^N C_h)$$

where M is a big positive value (in fact, it is enough if $M > 1$).

As noted in Kadziński and Tervonen (2013), when using a threshold-based sorting procedure, $C_N(a)$ is either empty or precise.

Case study Let us consider the same data set as in Table 1 with the aim of assigning 24 countries to 4 types of regimes: full democracies (C_4), flawed democracies (C_3), hybrid regimes (C_2), and authoritarian regimes (C_1). We assume a DM to have provided preference information in the form of the eight exemplary assignments (see Table 4).

The possible and necessary assignments are presented in Table 5. Obviously, the eight reference countries are necessarily assigned to the classes specified by the DM. Another nine non-reference alternatives are precisely assigned to a single class by all compatible value functions. For the remaining seven alternatives, the necessary assignment is empty and the possible assignments are imprecise. There are six countries possibly assigned to two consecutive classes (i.e., $C_1 - C_2$ or $C_2 - C_3$) and one country [Malayasia (MY)] with a possible assignment of three classes $C_2 - C_4$. The average width of the range of possible classes is 1.33.

4 Explaining recommendation in terms of decision maker's preference information

Justifying a recommendation to the DM by means of convincing arguments is a central concern for decision aiding tools (Labreuche et al. 2012). Generating a supportive evidence that would validate a certain choice, rank, or assignment is particularly difficult in case of multiple criteria preference models. Understanding the explanation that directly refers to the characteristics of these models may be simply too demanding for the non-experienced users for whom it is difficult to comprehend, e.g., the notions of criteria importance, compensation, or interactions between criteria. In particular, when referring to a single value function, exploiting it directly to justify the recommendation may not be meaningful to the DM (Labreuche 2011). In case of using a set of infinitely many compatible value functions as in ROR, even such direct exploitation of the preference model is not possible, because the DM can neither see a score of each alternative nor (s)he can assess relative importance of the criteria understood as a share of a given criterion in the comprehensive value.

Table 5 Possible ($C_L^P - C_R^P$) and necessary (C^N) assignments, all assignment preferential reducts (APRs) and the assignment preferential core (ACORE) for all countries (note that we refer to indices of assignment examples provided by the DM from Table 4)

Country	$C_L^P - C_R^P$	C^N	APRs	ACORE
AU	C4	C4	{1}, {2}	–
CN	C1	C1	{8}	{8}
FJ	$C2 - C3$	–	{3, 5, 7}, {3, 6, 7}	{3, 7}
ID	C3	C3	{3, 4, 5}, {3, 4, 6, 7, 8}	{3, 4}
JP	C4	C4	{1}	{1}
KZ	C2	C2	{6}	{6}
KG	C2	C2	{7}	{7}
LA	C1	C1	{8}	{8}
MY	$C2 - C4$	–	{6}, {7}	–
MN	$C2 - C3$	–	{3, 6}, {3, 7}	{3}
MM	C1	C1	{8}	{8}
NZ	C4	C4	{1}, {2}	–
KP	C1	C1	{1, 3, 4, 6, 8}, {1, 3, 5, 6, 8}	{1, 3, 6, 8}
PG	$C2 - C3$	–	{3, 5, 6}, {3, 5, 7}, {3, 7, 8}	{3}
PH	$C2 - C3$	–	{3, 7, 8}	{3, 7, 8}
SG	C3	C3	{5}	{5}
KR	C4	C4	{2}	{2}
TW	C3	C3	{3}	{3}
TJ	$C1 - C2$	–	{6, 7, 8}	{6, 7, 8}
TH	C3	C3	{4}	{4}
TL	C3	C3	{3, 4, 5, 6, 8}, {3, 4, 6, 7, 8}, {3, 4, 5, 7, 8}	{3, 4, 8}
TM	C1	C1	{8}	{8}
UZ	C1	C1	{5, 6, 8}	{5, 6, 8}
VN	$C1 - C2$	–	{6, 8}, {7, 8}	{8}

Traditionally, explanations that have been constructed to accompany the use of a value function in a decision aiding process referred to performances of some alternatives on different criteria and to importance of these criteria (see, e.g., [Greco et al. 2013](#); [Labreuche 2011](#); [Labreuche et al. 2012](#)). In this paper, we propose to emphasize the relation between the original preference information and a resulting recommendation. Explanations in our context will thus have the form of implications associating some pieces of DM's preference information with the necessary, possible, and extreme results. Precisely, we will refer to explanations of two kinds. On the one hand, the minimal set of preference information pieces, called preferential reduct, will be identified to justify some result that is observable for the set of compatible value functions, e.g., the truth of the necessary relation for some pair of alternatives. On the other hand, the maximal set of preference information pieces, called preferential construct, will be discovered to reveal the conditions under which some result non-observable for the set of compatible value functions, e.g., the falsity of the possible relation for some pair of alternatives, is possible. Provided with such explanations, the DM can better understand the impact of each piece of preference information on the result as well as can gain arguments in favor of a recommendation given by ROR in terms of some holistic judgments (s)he previously supplied. Such simple and intuitive explanations should either increase the DM's confidence in the result or stimulate her/his reactions in the

subsequent iterations in case (s)he finds some generated explanation (i.e., implication between the input and the output) unacceptable.

Let us remind that computation of the necessary, possible, and extreme results relies on solving a series of special optimization problems. Precisely, the analysis of infeasibility of these problems is of the utmost importance. For example, confirmation of the necessary preference relation or class assignment is based on the proof that constraints corresponding to the opposite preference relation or different assignment, respectively, are in conflict with the constraints defining the set of compatible value functions. Furthermore, when the latter constraints contradict the mathematical conditions corresponding to some preference relation or class assignment, such recommendation is considered as not possible for the set of compatible value functions. As a result, constructing an explanation of some recommendation provided by ROR amounts at identifying the reasons of infeasibility.

Since infeasibility is often encountered during the process of initial model formulation or reformulation, some algorithmic assistance is available to diagnose its cause. As noted in Chinneck (2008), applications of infeasibility analysis can be found, e.g., in backtracking in constraint logic programs, training neural networks, statistical analysis, video broadcasting, protein folding, radiation therapy, and signal processing. In these contexts, two major problems have been raised in the literature. The first one concerns isolating the infeasibility to a subset of constraints, called irreducible infeasible system (IIS) (see Carver 1921; Chinneck 1994; van Loon 1981).

Definition 1 Irreducible infeasible system is a subset of constraints defining the overall initial set of constraints (mathematical program) that itself is infeasible, but for which any proper subset is feasible; the IIS is minimal in the sense that it contains all constraints that contribute to the infeasibility.

Isolating an IIS from the larger model speeds the analysis, diagnosis and repair of the model by focusing the analytic effort to a small portion of the entire model. To emphasize the correspondence between IIS and explanations, we aim to construct, let us note that, e.g., in case of ranking problems, a preferential reduct of the necessary preference relation $a \succsim^N b$ is a small set of pairwise comparisons from among the many provided by the DM, so that in some way each of them contributes to the truth of $U(a) \geq U(b)$ for all compatible value functions, i.e., they all have to be there for infeasibility of $U(b) > U(a)$.

When dealing with the infeasibility, another problem is to determine the minimal set of constraints that have to be removed to restore the consistency in the initial system. It is known as solving the minimum IIS set-covering problem, or finding the maximum feasibility set (MFS) (see, e.g., Amaldi et al. 1999; Sankaran 1993).

Definition 2 Maximum feasibility set is a subset of constraints defining the overall initial set of constraints (mathematical program) that itself is feasible, but for which any proper superset is infeasible; the MFS is maximal in the sense that it contains all constraints that do not contradict each other.

Let us note that, e.g., in case of ranking problems, a preferential construct of the possible preference relation $a \succsim^P b$ for a pair (a, b) , such that $\text{not}(a \succsim^P b)$, is a set of

pairwise comparisons from among the many provided by the DM, so that all of them make $U(a) \geq U(b)$ possible, i.e., no additional pairwise comparison can be added to this set without contradicting $U(a) \geq U(b)$.

Algorithms for finding IISs and MFSs of constraints have been proposed, implemented, and tested in the recent years (for a review, see [Chinneck 2008](#)). In this section, we adapt these procedures to identify the preferential reducts and constructs. We discuss in detail the approaches for generating explanations of the necessary and possible preference relations. Analogously, we present the procedures for explaining the outcomes of ERA as well as the necessary and possible assignments, however, not discussing them in depth again. We also present preferential reducts and constructs for the case study concerning evaluation of democracy in the 24 countries.

4.1 Explaining recommendation of robust multiple criteria ranking

4.1.1 Necessary and possible preference relation

Let us first consider the problem concerning identification of the minimal set of pairwise comparisons provided by the DM that induces $a \succsim^N b$ in consequence of using all compatible value functions.

Definition 3 Assuming the use of a particular preference model, a preferential reduct for a pair $(a, b) \in A \times A$, such that $a \succsim^N b$, is a minimal set of pairwise comparisons provided by the DM inducing $a \succsim^N b$. Preferential reduct is minimal in a sense that any of its proper subsets does not imply $a \succsim^N b$. Let us denote it by $\text{PR}(a \succsim^N b)$.

Remark 1 Considering a pair $(a, b) \in A \times A$, such that $a \succsim^N b$ and a corresponding infeasible constraint set $E^N(a, b)$ with $\varepsilon > 0$, the following holds:

- Since E_{RANK}^A is feasible, IIS in $E^N(a, b)$ needs to contain constraint $U(b) - U(a) \geq \varepsilon$. Thus, IIS contains constraints on monotonicity and normalization of the preference model and/or constraints corresponding to pairwise comparisons provided by the DM that are in conflict with $U(b) > U(a)$, i.e., induce $U(a) \geq U(b)$ for all compatible value functions.
- A preferential reduct for $a \succsim^N b$ is equivalent to a set of pairwise comparisons whose corresponding constraints are included in an IIS in $E^N(a, b)$.

Let us now discuss a few algorithms for identification of a preferential reduct for $a \succsim^N b$. They are based on the procedures for isolating IIS in the set of infeasible constraints. Let us emphasize, however, that a preferential reduct for $a \succsim^N b$ is not equivalent to IIS in $E^N(a, b)$. In fact, we are only interested in the part of IIS that is useful in explaining the necessary relation $a \succsim^N b$ in terms of pairwise comparisons the DM previously supplied. We intentionally neglect the part that is related to the constraints on monotonicity and normalization. This implies the necessary adaptation of the algorithms for discovering IIS. In particular, the suitably adapted deletion filter ([Chinneck and Dravnieks 1991](#)) operates by testing the feasibility of the model when constraints corresponding to the provided pairwise comparisons are dropped one by one (see Algorithm 1). At the end of a single pass through all these constraints, the

algorithm identifies a single preferential reduct. Precisely, it indicates a preferential reduct whose first pairwise comparison makes the tested set of constraints feasible once eliminated.

Algorithm 1 Deletion filter for identifying a single preferential reduct for $a \succsim^N b$

Require: $E^N(a, b)$, where ε is an arbitrarily small positive value; it is an infeasible set of constraints

Ensure: PR, preferential reduct for $a \succsim^N b$.

```

1: PR =  $\emptyset$ .
2: for  $k \in \{1, \dots, |B^R|\}$  do
3:    $E^N(a, b) = E^N(a, b) \setminus B_k^R$ .
4:   Test the feasibility of  $E^N(a, b)$ .
5:   if feasible then
6:      $E^N(a, b) = E^N(a, b) \cup B_k^R$ .
7:     PR = PR  $\cup B_k^R$ .
8:   end if
9: end for

```

Example 1 Let us compute a preferential reduct for $(PG \succsim^N MY)$ using a deletion filter, where B^R is a set of pairwise comparisons provided by the DM and B_k^R , $k = 1, \dots, |B^R|$, are particular pairwise comparisons (see Table 2):

1. $\{B_2^R, B_3^R, B_4^R, B_5^R\}$ is feasible; PR = $\{B_1^R\}$ (B_1^R is added to PR).
2. $\{B_1^R, B_3^R, B_4^R, B_5^R\}$ is infeasible; PR = $\{B_1^R\}$ (B_2^R is not added to PR).
3. $\{B_1^R, B_4^R, B_5^R\}$ is feasible; PR = $\{B_1^R, B_3^R\}$ (B_3^R is added to PR).
4. $\{B_1^R, B_3^R, B_5^R\}$ is infeasible; PR = $\{B_1^R, B_3^R\}$ (B_4^R is not added to PR).
5. $\{B_1^R, B_3^R\}$ is feasible; PR = $\{B_1^R, B_3^R, B_5^R\}$ (B_5^R is added to PR).

Thus, a preferential reduct for $(PG \succsim^N MY)$ is equal to: $\text{PR}(PG \succsim^N MY) = \{B_1^R, B_3^R, B_5^R\} = \{JP \succ KR, PH \succ MN, KG \succ KZ\}$.

A single preferential reduct can be also identified with the suitably adapted additive method (Tamiz et al. 1996). This procedure adds in the constraints corresponding to the pairwise comparisons supplied by the DM, until infeasibility is achieved (see Algorithm 2). In this way, it indicates a preferential reduct whose last pairwise comparison makes the tested set of constraints infeasible once added.

Example 2 Let us compute a preferential reduct for $(PG \succsim^N MY)$ using the additive method.

1. $\{B_1^R\}$ is feasible, $\{B_1^R, B_2^R\}$ is feasible, $\{B_1^R, B_2^R, B_3^R\}$ is infeasible; PR = $\{B_3^R\}$; $\{B_3^R\}$ is feasible;
2. $\{B_1^R, B_3^R\}$ is feasible, $\{B_1^R, B_2^R, B_3^R\}$ is infeasible; PR = $\{B_2^R, B_3^R\}$; $\{B_2^R, B_3^R\}$ is infeasible.

Thus, a preferential reduct for $(PG \succsim^N MY)$ is equal to: $\text{PR}(PG \succsim^N MY) = \{B_2^R, B_3^R\} = \{FJ \succ TH, PH \succ MN\}$. Note that it is different from the one identified by the deletion filter.

Algorithm 2 Additive method for identifying a single preferential reduct for $a \succsim^N b$

Require: $E_{TS}^A = E_{BASE}^A \cup \{U(b) - U(a) \geq \varepsilon\}$, where ε is an arbitrarily small positive value.

Require: $E_{PI-RANK}^A$, a set of constraints corresponding to the provided pairwise comparisons.
 TS : the current test set of constraints.

Ensure: PR, preferential reduct for $a \succsim^N b$.

```

1: PR =  $\emptyset$ 
2: Test the feasibility of  $E_{TS}^A$ .
3: if infeasible then
4:   break
5: end if
6: TS =  $E_{TS}^A \cup$  PR.
7: for  $k \in \{1, \dots, |B^R|\}$  do
8:   TS = TS  $\cup B_k^R$ .
9:   Test the feasibility of TS.
10:  if infeasible then
11:    PR = PR  $\cup B_k^R$ .
12:    Go to 15.
13:  end if
14: end for
15: Test the feasibility of  $E_{TS}^A \cup$  PR.
16: if feasible then
17:   Go to 6.
18: end if
```

As noted by Chinneck (2008), a more efficient procedure consists in combining the additive method with the deletion filter. Precisely, the additive method is applied until first infeasible subset of constraints is found. Then, the deletion filter is used to identify a preferential reduct in this subset. Another idea for speeding up the search consists in grouping the constraints (see Chinneck 2008). Precisely, once deleting (adding) a group of B_k^R which causes feasibility (infeasibility), the deletion filter (the additive method) should backtrack and delete (add) singly.

Both the deletion filter and the additive method identify a single preferential reduct. However, there may exist several preferential reducts for relation $a \succsim^N b$. To enumerate all of them, we can use Algorithm 3. It extends the additive method by progressively adding all possible subsets of constraints corresponding to the provided pairwise comparisons, and eliminating from the test list the proper supersets of these constraint sets that already implied the infeasibility.

Knowing all preferential reducts for $a \succsim^N b$, we can identify the preferential core (see Definition 6).

Definition 4 Assuming the use of a particular preference model, the preferential core for a pair $(a, b) \in A \times A$, such that $a \succsim^N b$, is the intersection of all preferential reducts for this relation. Let us denote it by $PCORE(a \succsim^N b)$.

Remark 2 If the preferential core for a pair $(a, b) \in A \times A$, such that $a \succsim^N b$ is non-empty, elimination of any pairwise comparison contained in the core would result in the denial of $a \succsim^N b$.

Example 3 Let us compute the preferential core ($PCORE$) for $(PG \succsim^N MY)$. Since there are two preferential reducts for this pair of alternatives: $PR_1(PG \succsim^N MY) =$

Algorithm 3 Additive method for identifying all preferential reducts for $a \succsim^N b$

Require: $E_{TS}^{AR} = E_{BASE}^{AR} \cup \{U(b) - U(a) \geq \varepsilon\}$, where ε is an arbitrarily small positive value.

Require: $E_{PI-RANK}^{AR}$, a set of constraints corresponding to the provided pairwise comparisons.
 TS : the current test set of constraints.

Ensure: PRs, all preferential reducts for $a \succsim^N b$.

```

1: PRs =  $\emptyset$ 
2: Test the feasibility of  $E_{TS}^{AR}$ .
3: if infeasible then
4:   break
5: end if
6: SS = all subsets of  $E_{PI-RANK}^{AR}$  ordered with respect to the increasing cardinality.
7: for each  $SS_k \in SS$  do
8:   TS =  $E_{TS}^{AR} \cup SS_k$ .
9:   Test the feasibility of TS.
10:  if infeasible then
11:    PRs = PRs  $\cup SS_k$ .
12:    Remove all supersets of  $SS_k$  from SS.
13:  end if
14: end for

```

$\{B_2^R, B_3^R\}$ and $PR_2(PG \succsim^N MY) = \{B_1^R, B_3^R, B_5^R\}$, the preferential core is equal to $PCORE(PG \succsim^N MY) = \{B_3^R\} = \{PH \succ MN\}$.

The preferential cores for other pairs of alternatives can be computed as the intersection of all preferential reducts for a given pair provided in Fig. 1.

In case all preferential reducts are identified, we can also indicate a reduct with the fewest number of pairwise comparisons. Such a property is desirable for better understanding the reasons underlying the necessary relation and reducing the complexity of the subsequent analysis. However, as noted by Chakravarti (1994), the number of IISs in an infeasible linear programming (LP), and thus the number of preferential reducts for $a \succsim^N b$, could be exponential in the worst case. This motivated experiments aiming at indicating some heuristic methods discovering single IIS having as few constraints as possible. In this regard, Chinneck (1997) reports the elastic filter and deletion/sensitivity filter to be the most effective heuristics on a set of test problems.

Case study For the case study, all preferential reducts for the necessary relations visible in the Hasse diagram are provided in Fig. 1. Let us explicitly provide their interpretation for a few pairs of alternatives:

- $PR(MN \succsim^N SG) = \{B_3^R, B_5^R\} \Rightarrow$ “the necessary preference of MN over SG is a consequence of the DM’s two pairwise comparisons: $B_3^R = (PH \succ MN)$ and $B_5^R = (KG \succ KZ)$ ”;
- $PR_1(TW \succsim^N TL) = \{B_3^R\}$ and $PR_2(TW \succsim^N TL) = \{B_4^R\} \Rightarrow$ “the necessary preference of TW over TL is a consequence of the DM’s single pairwise comparison: either $B_3^R = (PH \succ MN)$ or $B_4^R = (ID \succ MY)$ ”.

In this way, the DM can precisely identify the reasons of the necessary relation’s truth. Note that for some pairs of alternatives, there is only a single preferential reduct [e.g., $(KR \succsim^N TL)$ and $(PG \succsim^N SG)$], while for the others there are more of them [e.g., $(KZ \succsim^N TJ)$ and $(VN \succsim^N UZ)$]. In case the preferential core is non-empty [e.g.,

$\text{PCORE}(\text{KR} \succsim^N \text{TL}) = \{B_5^R\}$ or $\text{PCORE}(\text{TL} \succsim^N \text{MY}) = \{B_3^R\}$, removal of any of its members from B^R would result in the falsity of the corresponding necessary relation.

The above-described algorithms may be easily adapted to explain the strict necessary preference $a \succ_U^N b \Leftrightarrow \forall U \in \mathcal{U}^R, U(a) > U(b)$ [we introduce notation \succ_U^N to avoid confusion with the previous papers on ROR, which interpret \succ^N as the asymmetric part of \succsim^N , i.e., $a \succ^N b \Leftrightarrow a \succsim^N b$ and not $(b \succsim^N a)$]. Precisely, in the deletion filter one should use $E^N(a, b) = E_{\text{RANK}}^{AR} \cup \{U(b) - U(a) \geq 0\}$, while in the additive method $E_{\text{TS}}^{AR} = E_{\text{BASE}}^{AR} \cup \{U(b) - U(a) \geq 0\}$. Thus, constraint $U(b) - U(a) \geq \varepsilon$ is replaced by $U(b) - U(a) \geq 0$. Moreover, since the following relations hold:

$$\begin{aligned} a \succ_U^N b &\Leftrightarrow \text{not}(b \succsim^P a) \Leftrightarrow \nexists U \in \mathcal{U}^R : U(b) \geq U(a), \\ a \succsim^N b &\Leftrightarrow \text{not}(b \succ_U^P a) \Leftrightarrow \nexists U \in \mathcal{U}^R : U(b) > U(a), \end{aligned}$$

the algorithms for identifying the preferential reducts for the necessary relation may also be used for indicating the minimal subset of pairwise comparisons implying that some relation is not even possible.

Let us now consider the problem concerning identification of the maximal set of pairwise comparisons provided by the DM that admits $a \succsim^P b$ in consequence of using at least one compatible value function, in case $\text{not}(a \succsim^P b)$, i.e., the problem of preferential constructs.

Definition 5 Assuming the use of a particular preference model, a preferential construct for a pair $(a, b) \in A \times A$, such that $\text{not}(a \succsim^P b)$, is a maximal set of pairwise comparisons provided by the DM admitting $a \succsim^P b$. Preferential construct is maximal in a sense that any of its proper supersets does not allow $a \succsim^P b$. Let us denote it by $\text{PC}(\text{not}(a \succsim^P b))$.

Remark 3 Considering a pair $(a, b) \in A \times A$, such that $\text{not}(a \succsim^P b)$ and a corresponding infeasible constraint set $E^P(a, b)$ with $\varepsilon > 0$, the following holds:

- Since E_{RANK}^{AR} is feasible, constraint $U(a) - U(b) \geq 0$ is in conflict with some constraints of E_{RANK}^{AR} , thus making $E^P(a, b)$ infeasible.
- Preferential construct for $\text{not}(a \succsim^P b)$ is equivalent to a set of pairwise comparisons whose corresponding constraints are included in a MFS together with constraint set E_{BASE}^{AR} and constraint $U(a) - U(b) \geq 0$.
- Identifying such a MFS is equivalent to finding a minimal set of pairwise comparisons whose corresponding constraints need to be removed from $E^P(a, b)$ so that the remaining constraint set is feasible.

Identification of a preferential construct for $\text{not}(a \succsim^P b)$ can be achieved by means of MILP (see, e.g., [Chinneck 2001](#); [Mousseau et al. 2003](#)). Using binary variables v_k associated with each pairwise comparison B_k^R provided by the DM, one rewrites the set of constraints $E^P(a, b)$ into $E_V^P(a, b)$ given below. If $v_k = 1$, then the corresponding constraint is always satisfied, which is equivalent to elimination of this constraint.

Thus, to identify a minimal subset of pairwise comparisons that need to be removed so that $U(a) \geq U(b)$ is possible, we solve the following problem:

$$\text{Maximize : } \sum_{k=1}^{|B^R|} v_k \quad (8)$$

s.t.

$$\left. \begin{array}{l} U(a) - U(b) \geq 0, \\ E_{\text{BASE}}^{A^R}, \\ v_k \in \{0, 1\}, \\ \text{if } a^* \succ_{\text{DM}} b^* \\ \quad Mv_k + U(a^*) \geq U(b^*) + \varepsilon, \\ \text{if } a^* \sim_{\text{DM}} b^* \\ \quad \left. \begin{array}{l} Mv_k + U(a^*) \geq U(b^*) \\ Mv_k + U(b^*) \geq U(a^*) \end{array} \right\} \\ \text{if } a^* \prec_{\text{DM}} b^* \\ \quad Mv_k + U(a^*) \geq U(b^*), \end{array} \right\} \left. \begin{array}{l} B_k^R, \quad k = 1, \dots, |B^R| \\ E_{\text{PI-RANK}}^{V, A^R} \end{array} \right\} E_V^P(a, b)$$

where M and ε are arbitrarily large and small positive values, respectively. As a result, the constraints corresponding to $v_k = 0$ indicate pairwise comparisons included in a preferential construct for $\text{not}(a \succsim^P b)$.

Other preferential constructs can be identified by adding constraints that forbid finding again the same solutions which have been already identified in the previously conducted optimizations:

$$\sum_{k \in S_i} v_k \leq f_i^* - 1,$$

where f_i^* is the optimal value of the objective function in the i th iteration, v_k^* are values of the binary variables at the optimum found while identifying the i th minimal subset standing behind incompatibility, and $S_i = \{k \in \{1, \dots, |B^R|\} : v_k^* = 1\}$.

Example 4 Let us compute all preferential constructs for a pair $(\text{SG}, \text{PH}) \in A \times A$, such that $\text{not}(\text{SG} \succsim^P \text{PH})$:

1. $S_1 = \{B_3^R\}$, and thus $\text{PC}_1(\text{not}(\text{SG} \succsim^P \text{PH})) = \{B_1^R, B_2^R, B_4^R, B_5^R\}$ (add $v_3 \leq 0$ to the set of constraints);
2. $S_2 = \{B_5^R\}$, and thus $\text{PC}_2(\text{not}(\text{SG} \succsim^P \text{PH})) = \{B_1^R, B_2^R, B_3^R, B_4^R\}$ (add $v_5 \leq 0$ to the set of constraints);
3. The set of constraints is infeasible, meaning that all preferential constructs have been already identified.

These results should be interpreted in the following way: “two sets of the DM’s pairwise comparisons: $\{B_1^R, B_2^R, B_4^R, B_5^R\}$ and $\{B_1^R, B_2^R, B_3^R, B_4^R\}$ make the preference of SG over PH still possible”, or from another perspective: “if the DM eliminated from B^R

either B_3^R or B_5^R , SG would be ranked at least as good as PH for at least one compatible value function”.

Analogously as for the case of reducts, it is interesting to consider the intersection of all preferential constructs.

Definition 6 Assuming the use of a particular preference model, the preferential core of constructs for a pair $(a, b) \in A \times A$, such that $\text{not}(a \succsim^P b)$, is the intersection of all preferential constructs for this relation. Let us denote it by $\text{PCCORE}(\text{not}(a \succsim^P b))$.

Remark 4 Elimination of all pairwise comparisons contained in the preferential core $\text{PCCORE}(\text{not}(a \succsim^P b))$ would not change the falsity of $a \succsim^P b$. From another perspective, these pairwise comparisons are not contained in any preferential reduct for a relation $b \succ_U^N a$.

Example 5 The preferential core of constructs for a pair $(\text{SG}, \text{PH}) \in A \times A$, such that $\text{not}(\text{SG} \succsim^P \text{PH})$ is $\text{PCCORE}(\text{not}(\text{SG} \succsim^P \text{PH})) = \{B_1^R, B_2^R, B_4^R\}$. The sole preferential reduct for a relation $\text{PH} \succ_U^N \text{SG}$ is $\{B_3^R, B_5^R\}$.

4.1.2 Extreme ranking analysis

Definition 7 Assuming the use of a particular preference model, a rank preferential reduct for an alternative $a \in A$, whose best and worst possible ranks are $P^*(a)$ and $P_*(a)$, respectively, is a minimal set of pairwise comparisons provided by the DM inducing a is neither ranked better than $P^*(a)$ nor worse than $P_*(a)$ by any compatible preference model instance. Let us denote it by $\text{RPR}(a)$.

A rank preferential reduct for $a \in A$, which guarantees a is ranked between $P^*(a)$ and $P_*(a)$ by all compatible value functions, can be identified using Algorithm 2, where $E_{\text{TS}}^{A^R}$ provided as the input is equal to $E^P(a \rightarrow [P^*(a), P_*(a)])$ given below.

$$\left. \begin{array}{l}
 \text{if } P^*(a) > 1 \\
 \left. \begin{array}{l}
 U(a) + Mv_b^> \geq U(b) + \varepsilon, \\
 \text{for } b \in A \setminus \{a\} \\
 \sum_{b \in A \setminus \{a\}} v_b^> \\
 \leq P^*(a) - 2 + |A| \cdot v_1, \\
 v_b^> \in \{0, 1\}, \text{ for } b \in A \setminus \{a\},
 \end{array} \right\} E^P(a \rightarrow < P^*(a)) \\
 \text{if } P_*(a) < |A| \\
 \left. \begin{array}{l}
 U(b) + Mv_b^< \leq U(a) + \varepsilon, \\
 \text{for } b \in A \setminus \{a\} \\
 \sum_{b \in A \setminus \{a\}} v_b^< \\
 \leq |A| - P_*(a) - 1 + |A| \cdot v_2, \\
 v_b^< \in \{0, 1\}, \text{ for } b \in A \setminus \{a\},
 \end{array} \right\} E^P(a \rightarrow > P_*(a)) \\
 v_1 + v_2 = 1, \\
 v_1, v_2 \in \{0, 1\}, \\
 E_{\text{RANK}}^{A^R}
 \end{array} \right\} E^P(a \rightarrow [P^*(a), P_*(a)])$$

Constraint set $E^P(a \rightarrow < P^*(a))$ guarantees a is ranked $(P^*(a) - 1)$ in the worst case. Constraint set $E^P(a \rightarrow > P_*(a))$ ensures a is ranked $(P_*(a) - 1)$ in the best case. Constraint $(v_1 + v_2 = 1)$ guarantees that either $E^P(a \rightarrow < P^*(a))$ or $E^P(a \rightarrow > P_*(a))$ is instantiated with $v_1 = 0$ or $v_2 = 0$, respectively, while the other constraint set is relaxed with $v_1 = 1$ or $v_2 = 1$, being always satisfied. Thus, finding a rank preferential reduct amounts at identifying a minimal set of pairwise comparisons which contradict both $E^P(a \rightarrow < P^*(a))$ and $E^P(a \rightarrow > P_*(a))$. Chinneck (2008) suggests that in case of MILP, one should rather use the additive/deletion filter. This allows removing by the deletion filter some potentially dubious constraints from the initial solution suggested by the additive method. The rank preferential core (*RCORE*) is defined analogously to *PCORE* (see Definition 6).

Case study For the case study, all rank preferential reducts are provided in Table 3. Their discussion should be enhanced by the analysis of the ERA in case there is no preference information (see $P_D^*(a) - P_*^D(a)$ in Table 3). Let us explicitly provide interpretation of these results for a few alternatives:

- $RPR(TW) = \{B_3^R\}$, $P^*(TW) = P_D^*(TW) = 3$ and $P_*(TW) = 5 < P_*^D(TW) = 12 \Rightarrow$ “TW is not ranked at a position worse than 5 (i.e., between 6 and 12) by any compatible value function, because the DM provided a pairwise comparison $B_3^R = (PH \succ MN)$ ”;
- $RPR(MN) = \{B_2^R, B_3^R, B_5^R\}$, $P^*(MN) = 6 > P_D^*(MN) = 4$ and $P_*(MN) = 9 < P_*^D(MN) = 14 \Rightarrow$ “MN is not ranked at a position better than 6 (i.e., between 4 and 5) nor worse than 9 (i.e., between 10 and 14) by any compatible value function, because the DM provided pairwise comparisons $B_2^R = (FJ \succ TH)$, $B_3^R = (PH \succ MN)$, and $B_5^R = (KG \succ KZ)$ ”.

This allows identification of the precise implications between the provided pairwise comparisons and the obtained ranking intervals. For the two countries (AU and NZ), there is not any rank preferential construct, which means that positions of these alternatives stem only from their evaluations and the assumed preference model (while the preference information does not influence the attained ranks). For the other 11 countries, *RPR* is unique, whereas for the remaining ones there are at least two *RPR*s. Again, in case *RCORE* is non-empty (e.g., $RCORE(FJ) = \{B_2^R, B_5^R\}$ or $RCORE(PG) = \{B_1^R\}$), elimination of any of its members from B^R would result in widening the ranking interval of the corresponding alternative.

Definition 8 Assuming the use of a particular preference model, an upper (lower) rank preferential construct for an alternative $a \in A$, whose best (worst) possible rank is $P^*(a)$ ($P_*(a)$), is a maximal set of pairwise comparisons provided by the DM admitting a is possibly ranked better than $P^*(a)$ (worse than $P_*(a)$). Let us denote it by $URPC(a)$ ($LRPC(a)$).

An upper rank preferential construct for $a \in A$, admitting a is possibly ranked better than $P^*(a)$, is identified by solving (8), where the constraint set $E_V^P(a, b)$ is replaced by $E^P(a \rightarrow < P^*(a))$ given below.

$$\left. \begin{array}{l} (C1) \ U(a) + Mv_b \geq U(b) + \varepsilon, \\ (C2) \ \sum_{b \in A \setminus \{a\}} v_b \leq P^*(a) - 2, \\ (C3) \ v_b \in \{0, 1\}, \quad \text{for } b \in A \setminus \{a\}, \\ E_{\text{BASE}}^{AR}, \\ E_{\text{PI-RANK}}^{V, AR}. \end{array} \right\} E^P(a \rightarrow < P^*(a))$$

Constraints (C1), (C2), and (C3) ensure that a is ranked not worse than $P^*(a) - 1$.

A lower rank preferential construct for $a \in A$, admitting a is possibly ranked worse than $P_*(a)$, is identified analogously with constraints (C1) and (C2) being replaced by (C4) and (C5) given below:

$$\left. \begin{array}{l} (C4) \ U(b) + Mv_b \geq U(a) + \varepsilon, \\ (C5) \ \sum_{b \in A \setminus \{a\}} v_b \leq |A| - P_*(a) - 1. \end{array} \right\}$$

Example 6 For an alternative TJ, whose extreme ranks are 17 and 22:

- There is a single upper rank preferential construct $\text{URPC}(\text{TJ}) = \{B_1^R, B_2^R, B_4^R, B_5^R\}$, which means that “ B_3^R is the only pairwise comparison in B^R preventing TJ from being ranked better than 17th”;
- There are two lower rank preferential constructs $\text{LRPC}_1(\text{TJ}) = \{B_1^R, B_3^R, B_4^R\}$ and $\text{LRPC}_2(\text{TJ}) = \{B_3^R, B_4^R, B_5^R\}$, which means that “either (B_2^R and B_5^R) or (B_1^R and B_2^R) need to be removed from the set of provided pairwise comparisons B^R to allow TJ being ranked worse than 22nd”.

4.2 Explaining recommendation of robust multiple criteria sorting

Definition 9 Assuming the use of a particular preference model, an assignment preferential reduct for an alternative $a \in A$, whose worst and best possible classes are $C_{LP}(a)$ and $C_{RP}(a)$, respectively, is a minimal set of assignment examples provided by the DM inducing a is never assigned to a class better than $C_{RP}(a)$ and worse than $C_{LP}(a)$. Let us denote it by $\text{APR}(a)$.

Remark 5 If $L = R$, a is necessarily assigned to class $C_{L=R}(a)$. Then, an assignment preferential reduct induces $a \rightarrow^N C_{L=R}(a)$.

An assignment preferential reduct for $a \in A$, which guarantees a is assigned to a class between $C_{LP}(a)$ and $C_{RP}(a)$ by all compatible value functions, can be identified using Algorithm 2, where E_{TS}^{AR} provided as the input is equal to $E(a \rightarrow^P [C_{LP}(a), C_{RP}(a)])$ given below.

$$\left. \begin{array}{l} \text{if } L_P > 1 : \\ (C6) \ U(a) + \varepsilon \leq t_{L_P-1} + Mv_1, \\ \text{if } R_P < p : \\ (C7) \ U(a) \geq t_{R_P} - Mv_2, \\ v_1 + v_2 = 1, \\ v_1, v_2 \in \{0, 1\}, \\ E_{\text{SORT}}^{AR}. \end{array} \right\} E\left(a \rightarrow^P [C_{LP}(a), C_{RP}(a)]\right)$$

Constraint (C6) guarantees that a is assigned to a class $C_{L_P(a)-1}$ or worse. Constraint (C7) ensures a is assigned to a class $C_{R_P(a)+1}$ or better. Constraint $(v_1 + v_2 = 1)$ guarantees that either (C6) or (C7) is instantiated with $v_1 = 0$ or $v_2 = 0$, respectively, ensuring either $U(a) < t_{L_P-1}$ or $U(a) \geq t_{R_P}$ is satisfied. Thus, finding an assignment preferential reduct is equivalent to identifying a minimal set of assignment examples which contradict both $U(a) < t_{L_P-1}$ and $U(a) \geq t_{R_P}$. The assignment preferential core (ACORE) is defined analogously to PCORE (see Definition 6).

Case study For the case study, all assignments preferential reducts and the assignment preferential cores are provided in Table 5. Let us explicitly provide interpretation of these results for a few alternatives:

- $\text{APR}_1(\text{AU}) = \{A_1^R\}$ and $\text{APR}_2(\text{AU}) = \{A_2^R\} \Rightarrow$ “the necessary assignment of AU to C_4 is a consequence of the DM’s single assignment example, either $A_1^R = (\text{JP} \rightarrow C_4)$ or $A_2^R = (\text{KR} \rightarrow C_4)$ ”;
- $\text{APR}(\text{SG}) = \{A_3^R, A_7^R, A_8^R\} \Rightarrow$ “SG is neither assigned to a class better than C_3 nor to a class worse than C_2 by any compatible value function, because the DM provided the following assignment examples: $A_3^R = (\text{TW} \rightarrow C_3)$, $A_5^R = (\text{SG} \rightarrow C_3)$, and $A_8^R = (\text{CN} \rightarrow C_1)$ ”.

This allows identification of the precise implications between the provided assignment examples and the obtained possible and necessary assignments. Note that if the range of possible assignments is a proper subset of the set of all classes (as it is the case for countries in the study), then APR is non-empty. The assignment examples contained in ACORE [e.g., $\text{ACORE}(\text{FJ}) = \{A_3^R, A_7^R\}$ or $\text{ACORE}(\text{MN}) = \{A_3^R\}$] need to be held in A^R to guarantee the possible assignment for an alternative is not less precise.

Definition 10 Assuming the use of a particular preference model, an upper (lower) assignment preferential construct for an alternative $a \in A$, whose best (worst) possible class is $C_{R_P(a)}$ ($C_{L_P(a)}$), is a maximal set of assignment examples provided by the DM admitting a is possibly assigned to a class better than $C_{R_P(a)}$ (worse than $C_{L_P(a)}$). Let us denote it by $\text{UAPC}(a)$ ($\text{LAPC}(a)$).

An upper assignment preferential construct for $a \in A$, which admits a is possibly assigned to a class better than $C_{R_P(a)}$, is identified by solving (8), where the constraint set $E_V^P(a, b)$ is replaced by $E(a \rightarrow^P \geq C_{R_P(a)+1})$ given below. Constraint $U(a) \geq t_{R_P}$ ensures that a is assigned to a class $C_{R_P(a)+1}$ or better.

$$\left. \begin{array}{l} U(a) \geq t_{R_P}, \\ E_{\text{BASE-SORT}}^{A^R}, \\ Mv_k + U(a^*) \geq t_{\text{LDM}(a^*)-1}, \\ -Mv_k + U(a^*) + \varepsilon \leq t_{\text{RDM}(a^*)}, \\ v_k \in \{0, 1\}. \end{array} \right\} A_k^R, \quad k = 1, \dots, |A^R| \quad E(a \rightarrow^P \geq C_{R_P(a)+1})$$

Example 7 There is a single upper assignment preferential construct for an alternative MN, whose best possible class is C_3 : $\text{UAPC}(\text{MN}) = \{A_1^R, A_2^R, A_4^R, A_5^R, A_6^R, A_7^R, A_8^R\}$, which means that “an exemplary assignment A_3^R is the only one in A^R that prevents MN from being possibly assigned to C_4 ”.

A lower assignment preferential construct for $a \in A$, which admits a is possibly assigned to a class worse than $C_{L_P(a)}$, is identified analogously with constraint $U(a) \geq t_{R_P}$ being replaced by $U(a) + \varepsilon \leq t_{L_P-1}$.

Example 8 There are two lower assignment preferential constructs for an alternative PH, whose worst possible class is C_2 : $LAPC_1(PH) = \{A_1^R, A_2^R, A_3^R, A_4^R, A_5^R, A_6^R, A_7^R\}$ and $LAPC_2(PH) = \{A_1^R, A_2^R, A_3^R, A_4^R, A_5^R, A_6^R, A_8^R\}$, which means that “either A_8^R or A_7^R need to be removed from the set of assignment examples A^R to allow PH being possibly assigned to C_1 ”.

Note that the notion of an upper (lower) assignment preferential construct is closely related to the previous work in the context of robust multiple criteria sorting with ELECTRE methods. In [Dias and Mousseau \(2003\)](#) and [Mousseau et al. \(2003\)](#), infeasibility analysis is applied to indicate assignment examples (or, in general, sets of constraints) which should be removed so as to assign an alternative to a better or worse class than, respectively, its best or worst possible class.

5 Computational cost

Let us denote the number of the provided preference information pieces by PI. For ranking problems, it is equal to the number of pairwise comparisons of reference alternatives $|B^R|$, while for sorting problems it is equal to the number of assignment examples $|A^R|$. To identify a single preferential reduct with a deletion filter or the additive method, in the worst case we need to solve, respectively, PI or $PI \cdot (PI - 1)/2$ problems. To discover all preferential reducts with an additive method, up to 2^{PI} problems need to be solved.

However, each of these problems is relatively small—the range of dimensions in case of using general monotonic marginal value functions is $n \cdot m + |B^R|$ constraints for preferential reducts, $n \cdot (m + 2) + |B^R|$ for rank preferential reducts, and $n \cdot m + 2|A^R|$ for assignment preferential reducts. The number of variables is equal to $n \cdot m$. In case of identifying rank preferential reducts, additional $2n$ binary variables are involved in problem formulation. Let us note that when using linear marginal functions the numbers of constraints and variables are significantly less. In particular, we use m variables instead of $n \cdot m$ ones.

Solving all these problems is not a burden for contemporary solvers. It is feasible in reasonably short time, even if all preferential reducts need to be discovered, because sets of reference alternatives in MCDA usually consist of modestly sized collections of choices. In fact, multiple criteria problems considered in operations research and management science (OR/MS) usually involve several dozens of alternatives ([Wallenius et al. 2008](#)). Thus, it is not realistic to assume that the DM would be willing to provide tens of preference statements when dealing with such a set of alternatives.

Obviously, in case of ranking problems, the number of pairwise comparisons may be great if the DM provides a complete pre-order for the set of reference alternatives, as in the original UTA method ([Jacquet-Lagreze and Siskos 1982](#)). Precisely, a ranking of $|A^R|$ alternatives can be decomposed to $|B^R| = |A^R| \cdot (|A^R| - 1)/2$ pairwise comparisons. For large sets A^R , indication of preferential reducts composed

of individual pairwise comparisons would be rather demanding in terms of required computation time. To reduce the computational complexity, we may treat all pairwise comparisons referring to a single reference alternative as a logical whole. Then, within the deletion filter, the corresponding constraints would be deleted in block, thus removing a single alternative from the reference ranking. Consequently, the solution indicated by a deletion filter would be a minimal complete pre-order of a subset of reference alternatives derived from the reference ranking provided by the DM. This complete pre-order may, however, contain some redundant elementary pairwise comparisons.

Let us note, however, that ROR was designed to deal with problems typically considered within MCDA, i.e., with problems concerning up to few hundred alternatives, and rather not thousands. For larger sets of alternatives, there would be too many optimization problems to compute, even for getting the basic outcomes of ROR, such as the necessary preference relation, extreme ranks, or possible assignment. Thus, ROR cannot be used to solve problems in information retrieval, natural language processing, or bio-informatics. Moreover, when the set of alternatives is very large, ROR is losing its advantage, because one can neither present to the DM all results (e.g., the ranking) and explanations for the alternatives, nor involve her/him in the interactive construction of preference information.

6 Using explanations in the interactive construction of preferences

Preferential reducts and constructs can be interpreted as transparent and easily understandable arguments that can be used to justify and explain the decision. They enable DM's understanding of the conditions for the necessary, possible, and extreme results by providing useful information on the role of particular preference information pieces in terms of consequences they imply. In the learning perspective of a constructive approach to decision aiding, these explanations can be a starting point for an interactive analysis and construction of DMs preferences. They may serve as a tool for looking more thoroughly into the subject, by reasoning, interpreting, and arguing. We permit the DM to confront her/his value system not only with results of applying the inferred model on the set of alternatives, but also with explanations that make her/his logic and assumptions explicit. This confrontation gives insight into her/his preferences, providing reactions in the subsequent iteration, as well as a better understanding of the employed method.

Preferential reducts and constructs are presented so as to invite the DM to an interaction. On the one hand, (s)he could not accept some explanation in case (s)he does not agree with the arguments it suggests. Let us suppose that the DM disagrees with a preferential reduct revealing the implication between a pair-wise comparison $a \succsim_{DM} b$ and the necessary preference relation $c \succsim^N d$, for $a, b, c, d \in A$. Then, (s)he can remove or invert a preference relation for a pair (a, b) . In the same spirit, in case the DM does not agree with the falsity of the possible preference relation $\text{not}(e \succsim^P f)$, $e, f \in A$, (s)he may wish to keep preference information pieces contained in its preferential construct only, and remove the remaining ones. On the other hand, the analysis of the provided explanations may also increase the DM's confidence in

some pieces of preference information. In case the DM agrees with the necessary or non-possible consequences, (s)he may confirm preference information contained in the corresponding preferential reduct or not contained in the preferential construct, respectively.

Analyzing the obtained results along with the explanations, the DM can judge if the suggested recommendation is logical and valid, and whether (s)he is satisfied with the correspondence between the preferences (s)he has at the moment and the recommendation based on the preference model. In this way, the process of preference construction is either continued (with strengthened preference information) or restarted (with revised preference information). The procedure ends when the DM finds acceptable the possible, necessary and extreme results, as well as the explanations given by preferential reducts and constructs.

7 Extensions

In this section, we discuss a few extensions of the basic procedures for computation of preferential reducts and constructs. For brevity, we focus on the necessary and possible preference relations, while an adaptation to ERA and multiple criteria sorting is omitted for being obvious.

Accounting for credibility of preference information ROR methods are designed to be used interactively with an incremental specification of preference information. Let $B_1^R \subseteq B_2^R \subseteq \dots \subseteq B_s^R$ be nested sets of DM's pairwise comparisons of reference alternatives associated with decreasing confidence levels. Each of those sets B_t^R , $t = 1, \dots, s$, is modeled with a set of constraints $E_{\text{RANK},t}^{AR}$ generating the set of compatible value functions \mathcal{U}_t^{AR} . Each time we pass from B_{t-1}^R to B_t^R , $t = 2, \dots, s$, we add to $E_{\text{RANK},t-1}^{AR}$ new constraints concerning pairs $(a, b) \in B_t^R \setminus B_{t-1}^R$. Thus, the sets of compatible value functions are nested in the inverse order of the related set of pairwise comparisons B_t^R , $t = 1, \dots, s$, i.e., $\mathcal{U}_1^{AR} \supseteq \mathcal{U}_2^{AR} \supseteq \dots \supseteq \mathcal{U}_s^{AR}$. We suppose that $\mathcal{U}_s^{AR} \neq \emptyset$. On the basis of nested sets of pairwise comparisons $B_1^R \subseteq B_2^R \subseteq \dots \subseteq B_s^R$, we can build nested necessary $\succsim_1^N \subseteq \succsim_2^N \subseteq \dots \subseteq \succsim_s^N$ and possible $\succsim_1^P \supseteq \succsim_2^P \supseteq \dots \supseteq \succsim_s^P$ preference relations.

The DM may require that the constructed preferential reducts and constructs contain the most credible preference information which is provided in the initial phases of the interactive process. Computation of such a preferential reduct for $a \succsim^N b$ requires identification of the first nested necessary relation \succsim_t^N , $t = 1, \dots, s$, containing $a \succsim^N b$. Then, to indicate the preferential reduct consisting of the most credible preference information, the adequate procedures should exploit a set of pairwise comparisons B_t^R rather than B_s^R . Analogously, computation of the most credible preferential construct for $a \succsim^P b$ requires identification of the last nested possible relation \succsim_t^P , $t = s, \dots, 1$, containing $a \succsim^P b$. Then, the adequate algorithm should exploit the set of pairwise comparisons $B_s^R \setminus B_t^R$ to eliminate the least credible statements making $a \succsim^P b$ impossible. Note that analogous idea has been explored before in case of dealing with inconsistency of preference information. Mousseau et al. (2006) introduce algorithms for providing solutions with the least confident judgments or constraints

to be relaxed/deleted with a higher priority than solutions recommending relaxation of high-confidence statements.

Joint reducts and constructs for multiple preference relations Apart from analyzing preferential reducts and constructs for a single preference relation, we can also consider analogous explanations in terms of a set of multiple relations, i.e., $NEC = \{a_1 \succsim^N b_1, a_2 \succsim^N b_2, \dots\}$ and $POS = \{a_1 \succsim^P b_1, a_2 \succsim^P b_2, \dots\}$, respectively. To identify a joint preferential reduct for NEC , within a deletion filter we should consider the following set of constraints:

$$\left. \begin{array}{l} M(1 - v_{a_i, b_i}) + U(b_i) - U(a_i) \geq \varepsilon, \quad \text{for each } (a_i \succsim^N b_i) \in NEC \\ \sum_{i=1}^{|NEC|} v_{a_i, b_i} = 1, \\ v_{a_i, b_i} \in \{0, 1\}, \\ E_{RANK}^{AR}, \end{array} \right\} E^N(NEC) \quad \text{for each } (a_i \succsim^N b_i) \in NEC,$$

instead of considering the set of constraints $E^N(a, b)$. Analogously, to compute a joint preferential construct for POS , within the set of constraints $E_V^P(a, b)$, we should replace a constraint $U(a) - U(b) \geq 0$ with a constraint set $U(a_i) - U(b_i) \geq 0$, for each $(a_i \succsim^P b_i) \in POS$.

Forcing explanations to contain some pairwise comparisons The DM may wish to obtain explanations referring to some background preference information, i.e., to get a reduct which contains some specific pairwise comparisons. This can be achieved by the rearrangement of the list of pairwise comparisons to be dropped one by one in the deletion filter, so that the constraints that should be potentially included in the reduct are tested first. Obviously, it may turn out that the indicated pieces of preference information are redundant with respect to explaining the investigated result, and thus they are not included in the preferential reduct. In this case, the DM should be informed that it is impossible to formulate a minimal explanation using the indicated arguments.

Necessary projection of a set of pairwise comparisons of reference alternatives Let us call the set of necessary preference relations induced by some subset of pairwise comparisons its necessary projection on the set of pairs of alternatives. Analyzing such a projection may be useful for the DM interested in the necessary relations, (s)he needs to accept as a consequence of some specific subset of pairwise comparisons of reference alternatives. In particular, when considering the projection of the preferential reduct for $a \succsim^N b$, the DM is provided with all other necessary relations induced by the same subset of preference information pieces and (s)he may judge whether their joint induction is concordant with her/his value system.

8 Conclusions

In this paper, we presented a new approach to generating explanations of the recommendations offered by multiple criteria ranking or sorting methods based on robust ordinal regression. We have suitably adapted algorithms for identification of irreducible infeasible systems and maximum feasibility sets to indicate preferential reducts and constructs, respectively. The previous associates some result that is valid within the set of compatible value functions with the minimal set of preference information

pieces by which it is supported. The latter indicates the maximal set of the holistic judgments which make some currently non-observable result possible for at least one compatible value function. Knowing the exact relations between the provided preference information and the final recommendation, the DM can better understand the impact of each piece of preference information or their subsets on the result, and gain a better insight into her/his preferences. This enhances the interpretability of the outcomes and increases the DM's acceptance of the suggested recommendation.

The presented approach remains valid for different types of holistic preference information. In particular, in case of ranking problems, apart from pairwise comparisons we may use the intensity of preference (e.g., a is preferred to b at least as much as c is preferred to d) (Figueira et al. 2009) or rank-related requirements (e.g., a should be ranked in top 3, b should be among the 10 worst alternatives, or c should be placed in second five) (Kadzinski et al. 2013). Moreover, suitably adapted, the proposed algorithms can be used in the context of other preference models, e.g., outranking-based ranking methods (Greco et al. 2011; Kadzinski et al. 2012) or fuzzy integrals (Angilella et al. 2010), as well as in the case of group decision (Greco et al. 2012; Kadzinski et al. 2013).

Acknowledgments The first author acknowledges support from the Polish National Science Center (Grant SONATA). The fourth author acknowledges financial support from the Polish National Science Center, Grant No. 155534.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Amaldi E, Pfetsch ME, Trotter LE, Jr (1999) Some structural and algorithmic properties of the maximum feasible subsystem problem. In: Proceedings of the 7th international IPCO conference on integer programming and combinatorial optimization. Springer, London, pp 45–59
- Amgoud L, Prade J (2009) Using arguments for making and explaining decisions. *Artif Intell* 173(34):413–436
- Angilella S, Greco S, Matarazzo B (2010) Non-additive robust ordinal regression: a multiple criteria decision model based on the Choquet integral. *Eur J Oper Res* 201(1):277–288
- Belton V, Branke J, Eskelinen P, Greco S, Molina J, Ruiz F, Słowiński R (2008) Interactive multiobjective optimization from a learning perspective. In: Branke J, Deb K, Miettinen K, Słowiński R (eds) *Multiobjective Optim.* Springer, Berlin, pp 405–433
- Carenini G, Moore JD (2006) Generating and evaluating evaluative arguments. *Artif Intell* 170(11):925–952
- Carver WB (1921) Systems of linear inequalities. *Ann Math* 23:212–220
- Chakravarti N (1994) Some results concerning post-infeasibility analysis. *Eur J Oper Res* 73(1):139–143
- Chinneck JW (1994) MINOS(IIS): infeasibility analysis using MINOS. *Comput Oper Res* 21(1):1–9
- Chinneck JW (1997) Finding a useful subset of constraints for analysis in an infeasible linear program. *INFORMS J Comput* 9(2):164–174
- Chinneck JW (2001) Fast heuristics for the maximum feasible subsystem problem. *INFORMS J Comput* 13(3):210–223
- Chinneck JW (2008) *Feasibility and infeasibility in optimization: algorithms and computational methods.* Springer, New York
- Chinneck JW, Dravnieks EW (1991) Locating minimal infeasible constraint sets in linear programs. *INFORMS J Comput* 3(2):157–168

- Corrente S, Greco S, Kadziński M, Słowiński R (2013) Robust ordinal regression in preference learning and ranking. *Mach Learn* 93(2–3):381–422
- Dias LC, Mousseau V (2003) IRIS: a DSS for multiple criteria sorting problems. *J Multi-Criteria Decis Anal* 12(1):285–298
- Doumpos M (2012) Learning non-monotonic additive value functions for multicriteria decision making. *OR Spectr* 34(1):89–106
- EIU, Democracy Index (2010) Democracy in retreat. Economist Intelligence Unit, London
- Figueira J, Greco S, Słowiński R (2009) Building a set of additive value functions representing a reference preorder and intensities of preference: GRIP method. *Eur J Oper Res* 195(2):460–486
- Fürnkranz J, Hüllermeier E (2010) Preference Learning. Springer, Berlin
- Greco S, Kadziński M, Mousseau V, Słowiński R (2011) ELECTRE^{GKMS}: robust ordinal regression for outranking methods. *Eur J Oper Res* 214(1):118–135
- Greco S, Kadziński M, Mousseau V, Słowiński R (2012) Robust ordinal regression for multiple criteria group decision problems: UTA^{GMS}-GROUP and UTADIS^{GMS}-GROUP. *Decis Support Syst* 52(3):549–561
- Greco S, Matarazzo B, Słowiński R (2004) Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *Eur J Oper Res* 158(2):271–292
- Greco S, Mousseau V, Słowiński R (2008) Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *Eur J Oper Res* 191(2):415–435
- Greco S, Mousseau V, Słowiński R (2010) Multiple criteria sorting with a set of additive value functions. *Eur J Oper Res* 207(4):1455–1470
- Greco S, Słowiński R, Zielniewicz P (2013) Putting dominance-based rough set approach and robust ordinal regression together. *Decis Support Syst* 54(2):891–903
- Greer JE, Falk S, Greer KJ, Bentham MJ (1994) Explaining and justifying recommendations in an agriculture decision support system. *Comput Electron Agric* 11(23):195–214
- Hammond JS, Keeney RL, Raiffa H (1998) Even Swaps: a rational method for making trade-offs. *Harvard Bus Rev* 76(2):137–138, 143–148, 150
- Humphreys P, Wisudha A (1981) MAUD: an interactive computer program for the structuring, decomposition, and recombination of preferences between multiattributed alternatives. Technical Report 543, US Army Research Institute for the Behavioral and Social Sciences, Alexandria
- Jacquet-Lagrèze E, Siskos Y (1982) Assessing a set of additive utility functions for multicriteria decision making: the UTA method. *Eur J Oper Res* 10(2):151–164
- Jacquet-Lagrèze E, Siskos Y (2001) Preference disaggregation: 20 years of MCDA experience. *Eur J Oper Res* 130(2):233–245
- Kadziński M, Greco S, Słowiński R (2012) Extreme ranking analysis in robust ordinal regression. *Omega* 40(4):488–501
- Kadziński M, Greco S, Słowiński R (2013) RUTA: a framework for assessing and selecting additive value functions on the basis of rank related requirements. *Omega* 41(4):735–751
- Kadziński M, Greco S, Słowiński R (2013) Selection of a representative value function for robust ordinal regression in group decision making. *Group Decis Negot* 22(3):429–462
- Kadziński M, Tervonen T (2013) Robust multi-criteria ranking with additive value models and holistic pair-wise preference statements. *Eur J Oper Res* 228(1):169–180
- Kadziński M, Tervonen T (2013) Stochastic ordinal regression for multiple criteria sorting problems. *Decis Support Syst* 55(1):55–66
- Kass R, Finin T (October 1988) The need for user models in generating expert system explanation. *Int J Expert Syst* 1(4):345–375
- Klein DA (1994) Decision analytic intelligent systems: automated explanation and knowledge acquisition. Lawrence Erlbaum Associates, Hillsdale, New Jersey
- Labreuche C (2011) A general framework for explaining the results of a multi-attribute preference model. *Artif Intell* 175:1410–1448
- Labreuche C, Maudet N, Mousseau V, Ouerdane W (2012) Explanation of the robust additive preference model by even swap sequences. In: 6th multidisciplinary workshop on advances in preference handling, pp 21–26
- Labreuche C, Maudet N, Ouerdane W (2011) Minimal and complete explanations for critical multi-attribute decisions. In: ADT, pp 121–134
- Lahdelma R, Salminen P (2012) The shape of the utility or value function in stochastic multicriteria acceptability analysis. *OR Spectr* 34(4):785–802

- Larichev O (2001) Ranking multicriteria alternatives: the method ZAPROS III. *Eur J Oper Res* 131(3):550–558
- March JG (1978) Bounded rationality, ambiguity and the engineering of choice. *Bell J Econ* 9:587–608
- Mousseau V, Dias LC, Figueira J (2006) Dealing with inconsistent judgments in multiple criteria sorting models. *4OR* 4(3):145–158
- Mousseau V, Figueira J, Dias LC, Gomes da Silve C, Climaco J (2003) Resolving inconsistencies among constraints on the parameters of an MCDA model. *Eur J Oper Res* 147(1):72–93
- Ouerdane W, Muadet N, Tsoukiás A (2010) Argumentation Theory and Decision Aiding. In: Figueira J, Greco S, Ehrgott M (eds) *International series in operations research and management science* 1, vol 142. *Trends in Multiple Criteria Decision Analysis*, pp 177–208
- Papamichail KN, French S (2003) Explaining and justifying the advice of a decision support system: a natural language generation approach. *Expert Syst Appl* 24(1):35–48
- Sankaran J (1993) A note on resolving infeasibility in linear programs by constraint relaxation. *Oper Res Lett* 13(1):19–20
- Słowiński R, Greco S, Matarazzo B (2009) Rough sets in decision making. In: Meyers RA (ed) *Encyclopedia of complexity and systems science*. Springer, New York, pp 7753–7787
- Tamiz M, Mardle SJ, Jones DF (1996) Detecting IIS in infeasible linear programmes using techniques from goal programming. *Comput Oper Res* 23(2):113–119
- van Loon JNM (1981) Irreducibly inconsistent systems of linear inequalities. *Eur J Oper Res* 8(3):283–288
- Wallenius J, Dyer JS, Fishburn PC, Steuer RE, Zionts S, Deb K (2008) Multiple criteria decision making, multiattribute utility theory: recent accomplishments and what lies ahead. *Manag Sci* 54(7):1336–1349